



## Strathprints Institutional Repository

Ren, Jinchang and Zabalza, Jaime and Marshall, Stephen and Zheng, Jiangbin (2014) *Effective feature extraction and data reduction with hyperspectral imaging in remote sensing*. IEEE Signal Processing Magazine, 31 (4). pp. 149-154. ISSN 1053-5888

Strathprints is designed to allow users to access the research output of the University of Strathclyde. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://strathprints.strath.ac.uk/>) and the content of this paper for research or study, educational, or not-for-profit purposes without prior permission or charge.

Any correspondence concerning this service should be sent to Strathprints administrator: <mailto:strathprints@strath.ac.uk>

# Effective Feature Extraction and Data Reduction in Remote Sensing Using Hyperspectral Imaging

Jinchang Ren, Jaime Zabalza, Stephen Marshall, Jiangbin Zheng

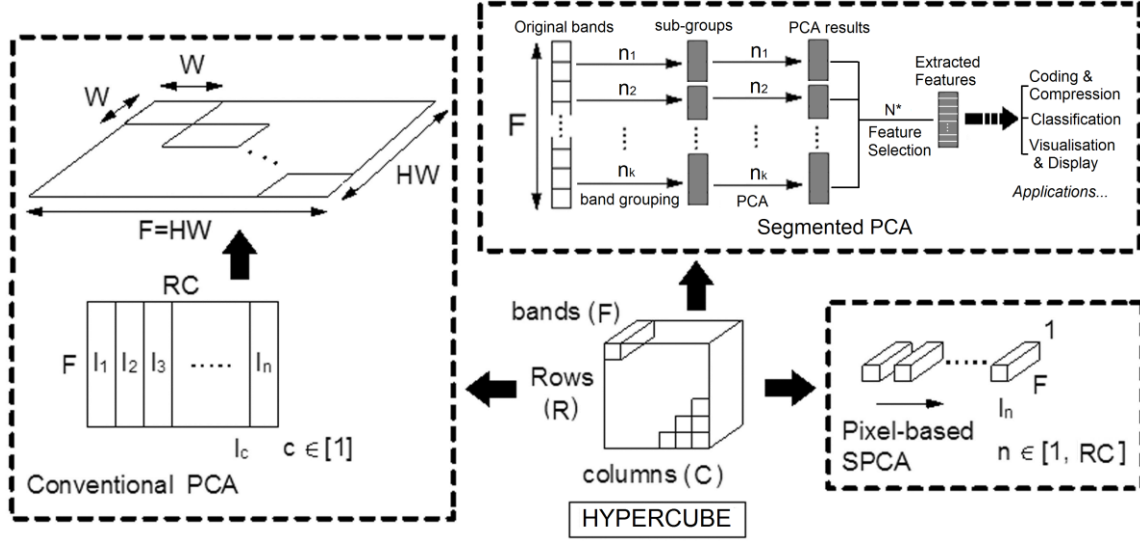
With numerous and contiguous spectral bands acquired from visible light (400nm-1000nm) to (near) infrared (1000nm-1700nm and over), hyperspectral imaging (HSI) can potentially identify different objects by detecting minor changes in temperature, moisture and chemical content. As a result, HSI has been widely applied in a number of application areas, including remote sensing [1]. HSI data contains 2-D spatial and 1-D spectral information, and naturally forms a 3-D hypercube with a high spectral resolution in nanometers that enables robust discrimination of ground features. However, new challenges arise in dealing with extremely large datasets. For a hypercube with relatively small spatial dimension of  $600 \times 400$  pixels at 16 bits-per-band-per-pixel, the data volume becomes 120M bytes for 250 spectral bands. In some cases, this large data volume can be linearly increased when multiple hypercubes are acquired across time to monitor system dynamics in consecutive time instants. When the ratio between the feature dimension (spectral bands) and the number of data samples (in vector based pixels) is vastly different, high dimensional data suffers from the well-known curse of dimensionality. For feature extraction and dimensionality reduction, Principal Components Analysis (PCA) is widely used in HSI [2], where the number of extracted components is significantly reduced compared to the original feature dimension, i.e. the number of spectral bands. For effective analysis of large scale data in HSI, conventional PCA faces three main challenges: (i) To obtain the covariance matrix in extremely large spatial dimension, which can lead to software tools such as Matlab running out of memory; (ii) To cope with the high computational cost required for analysis of large datasets; and (iii) To retain locally structured elements which only appear in a small number of bands (namely local structures) for improved discriminating ability when feature bands are globally extracted as principal components. This article discusses several variations and extensions of conventional PCA to address the aforementioned challenges. These variations and extensions include slicing the HSI data for efficient computation of the covariance matrix similarly done in 2D-PCA analysis [3], and grouping the spectral data to preserve the local structures and further speed-up the process to determine the covariance matrix [4]. In addition, we also discuss some non-PCA based approaches for feature extraction and data reduction, based on techniques such as band selection, random projection, singular value decomposition and machine learning approaches such as support vector machine (SVM) [2, 5, 6].

## FEATURE EXTRACTION USING PCA AND ITS VARIATIONS

PCA has been widely used for unsupervised feature extraction and data reduction [2, 7]. Through orthogonal projection and truncation of the transformed feature data, PCA can successfully remove correlation inherent in the data.

For a hypercube with  $F$  spectral bands and a spatial size of  $R \times C$  (Fig. 1), where  $R$  and  $C$  are the number of rows and columns, respectively, conventional PCA first converts the data into an  $F \times S$  matrix ( $\mathbf{I}$ ), where  $S=RC$ . Then, the covariance matrix of  $\mathbf{I}$ ,  $\mathbf{\Lambda}$ , is obtained as follows:

Eigen-decomposition is followed by data projection to determine the principal components. Due to the extremely large size of  $S$ , which can be over 100k bytes, practical difficulties arise when directly determining  $\Lambda$  from  $\mathbf{I}$ . To solve these problems, 2D-PCA inspired *data slicing PCA* [3] along with segmented PCA (Seg-PCA) [4] are used.



[Fig1] Diagrams of conventional PCA in comparison with the data slicing PCA (SPCA) and Segmented PCA (Seg-PCA)

### Data Slicing PCA (SPCA)

Rather than taking all spectral vectors in  $\mathbf{I}$  together, SPCA (Fig. 1) treats each spectral vector  $\mathbf{I}_n$  separately when determining the corresponding covariance matrix  $\Lambda$ , where  $n \in [1, S]$ . For each  $\mathbf{I}_n \in \mathbb{R}^{F \times 1}$ , its associated partial covariance matrix is obtained as  $\mathbf{I}_n \mathbf{I}_n^T \in \mathbb{R}^{F \times F}$ , and the overall covariance matrix  $\Lambda$  is determined as the sum of all partial covariance matrices by

$$\Lambda = \frac{1}{S} \sum_{n=1}^{S=RC} \mathbf{I}_n \mathbf{I}_n^T \quad (1)$$

SPCA is equivalent to PCA in determining the covariance matrix. The fundamental difference is that the calculation of  $\Lambda$  is implemented by a series of  $S$  independent partial covariance matrices. As such, the memory requirement is reduced from  $F \times S$  to  $F \times F$ . In addition to this, these partial covariance matrices can be separately calculated in parallel to improve efficiency.

### Segmented PCA (Seg-PCA)

Although the spectral data in HSI naturally form long vectors, the input spectral data can be grouped to produce a series of small vectors for fast calculation of PCA in each group whilst enabling extraction of local structures from each group. As shown in Fig 1, in Seg-PCA each of the  $S$  spectral vectors is grouped to form  $k$  sub-vectors. For each sub-vector, the size of its covariance matrix is reduced to  $n_k \times n_k$ .

For a particular group  $b \in [1, k]$ , let  $\mathbf{i}_{mb} = [i_{1m} \ i_{2m} \ \dots \ i_{n_b, m}]_b^T$  denote the corresponding small spectral vector formed from the original  $F$  bands, where  $m \in [1, S]$ , and  $\Lambda_b$  is the covariance matrix.

Then,  $\Lambda_b$  can be determined as follows.

$$\Lambda_b = \frac{1}{S} \sum_{m=1}^S \mathbf{i}_{mb} \mathbf{i}_{mb}^T \quad (2)$$

where again  $S$  denotes the spatial size of the hypercube. Also note that  $\mathbf{i}_{mb} \mathbf{i}_{mb}^T$  is the partial covariance matrix obtained from the grouped spectral vector  $\mathbf{i}_{mb}$ .

For a given spectral vector  $\mathbf{i}_n \in \mathbb{R}^{F \times 1}$ , conventional PCA extracts all the principal components from the  $F$  bands. In contrast, Seg-PCA divides  $\mathbf{i}_n$  into  $H$  groups and ensures that variable numbers of components are extracted from each group. As a result, Seg-PCA has the potential to preserve some local spectral structures, which are non-dominant and thus are discarded in conventional PCA. Such local structures provide additional values to Seg-PCA for superior discrimination ability as explained below.

In Seg-PCA, as illustrated in Fig. 1, each spectral vector of  $F$  bands is grouped into  $H$  groups. When  $H = 1$ , Seg-PCA defaults to SPCA, which is equivalent to the conventional PCA. In fact, the performance of the Seg-PCA is strongly dependent on how the bands are grouped, as it affects how much additional information can be extracted to improve the conventional PCA.

In one particular case where all groups contain the same number of  $W$  bands,  $HW = F$ , the covariance matrices from different groups share the same dimension. As a result, they can be summed to form one covariance matrix  $\Lambda_{Seg-PCA}$ . This can further simplify the eigen-decomposition procedure for better efficiency, as only one matrix rather than  $k$  must be processed.

$$\Lambda_{Seg-PCA} = \sum_{b=1}^k \Lambda_b \quad (3)$$

In conventional PCA, using SPCA for example, the covariance matrix in (1) can be rewritten as follows, where  $\mathbf{i}_n = [\mathbf{a}_{1n}, \mathbf{a}_{2n}, \dots, \mathbf{a}_{Hn}]$ .

$$\Lambda_{SPCA} = \frac{1}{S} \sum_{n=1}^S \mathbf{i}_n \mathbf{i}_n^T = \frac{1}{S} \sum_{n=1}^S \begin{bmatrix} \mathbf{a}_{1n} \mathbf{a}_{1n}^T & \mathbf{a}_{1n} \mathbf{a}_{2n}^T & \cdots & \mathbf{a}_{1n} \mathbf{a}_{Hn}^T \\ \mathbf{a}_{2n} \mathbf{a}_{1n}^T & \mathbf{a}_{2n} \mathbf{a}_{2n}^T & \cdots & \mathbf{a}_{2n} \mathbf{a}_{Hn}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{Hn} \mathbf{a}_{1n}^T & \mathbf{a}_{Hn} \mathbf{a}_{2n}^T & \cdots & \mathbf{a}_{Hn} \mathbf{a}_{Hn}^T \end{bmatrix} \quad (4)$$

For the particular case in Seg-PCA where all groups have the same number of bands, the covariance matrix can be further derived as:

$$\Lambda_{Seg-PCA} = \frac{1}{S} \sum_{n=1}^S \mathbf{I}_n \mathbf{I}_n^T = \frac{1}{S} \sum_{n=1}^S (\mathbf{a}_{1n} \mathbf{a}_{1n}^T + \mathbf{a}_{2n} \mathbf{a}_{2n}^T + \cdots + \mathbf{a}_{Hn} \mathbf{a}_{Hn}^T) \quad (5)$$

From (4) and (5), it is straightforward to demonstrate that  $\Lambda_{Seg-PCA}$  is formed by accumulating the  $W \times W$  sections in the main diagonal of  $\Lambda_{SPCA}$ , the original covariance matrix. This is also illustrated in Fig. 1 for comparison. Note that in a real hypercube, uneven band groupings are used to allow variable numbers of bands to be contained in different groups. However, this leads to the challenging problem of grouping bands for feature characterization, which is discussed in the following sections.

## FEATURE EXTRACTION USING NON-PCA BASED APPROACHES

For non-PCA based feature extraction and dimensionality reduction, the most commonly used approaches are based on band selection [5], machine learning [6] and steepest ascent search [7]. As adjacent spectral bands in hyperspectral imaging usually contain a large degree of redundant information, these approaches tend to select individual spectral bands rather than PCA components for data classification.

In these approaches, bands that contain less discriminatory information are discarded, which results in a much reduced number of remaining bands. Usually, the discriminatory information is reflected by how different a band is in comparison to adjacent bands. To determine the degree of similarity between two (band) images, a number of approaches can be used including distance measurement, mutual information and the structural similarity measurement (SSIM) [2, 5].

For any two band images  $A_m$  and  $A_n$ , the simplest distance-based similarity is given below to measure an average distance/similarity over all pixels:

$$S_d(A_m, A_n) = 1 - \frac{\sum_{i=1}^X \sum_{j=1}^Y |A_m(i, j) - A_n(i, j)|^\lambda}{\sum_{i=1}^X \sum_{j=1}^Y [\max(A_m(i, j), A_n(i, j))]^\lambda} \quad (6)$$

where  $\lambda > 0$  and  $(i, j)$  are the spatial coordinate indices of the two images satisfying  $i \in [1, X]$  and  $j \in [1, Y]$ ;  $X \times Y$  denotes the spatial dimension of the band images.

In contrast, mutual information based similarity is determined by the image histograms [2, 5]. By treating the spectral images as random variables, their associated mutual information can be determined as follows:

$$I(A_m, A_n) = H(A_m) + H(A_n) - H(A_m, A_n) \quad (7)$$

where  $H(A_m)$  and  $H(A_n)$  are the entropy of  $A_m$  and  $A_n$ , and  $H(A_m, A_n)$  is their joint entropy.

$$H(A_m) = - \sum_{a \in A_m} p(a) \log p(a) \quad (8)$$

The structural similarity measurement (SSIM) [8] contains three factors, including consistency in terms of luminance  $l(A_m, A_n)$ , contrast  $c(A_m, A_n)$  and structure  $s(A_m, A_n)$  defined as follows.

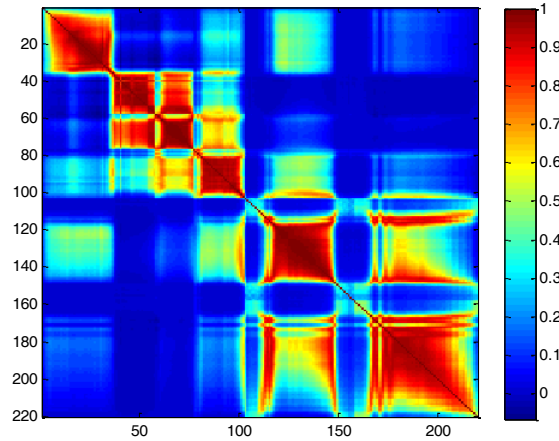
$$l(A_m, A_n) = \frac{2\mu_m\mu_n + C_1}{\mu_m^2 + \mu_n^2 + C_1} \quad c(A_m, A_n) = \frac{2\sigma_m\sigma_n + C_2}{\sigma_m^2 + \sigma_n^2 + C_2} \quad s(A_m, A_n) = \frac{\sigma_{mn} + C_3}{\sigma_m\sigma_n + C_3} \quad (9)$$

$$SSIM(A_m, A_n) = l^\alpha(A_m, A_n) c^\beta(A_m, A_n) s^\gamma(A_m, A_n) \quad (10)$$

where  $(\mu_m, \mu_n)$  and  $(\sigma_m, \sigma_n)$  are the mean intensity and the standard deviation of the band images  $A_m$  and  $A_n$ , respectively;  $C_1 - C_3$  are constants; and  $\alpha, \beta, \gamma$  are non-negative weights. In the particular case where  $\alpha = \beta = \gamma = 1$  and  $C_2 = 2C_3$ , SSIM can be further simplified as

$$SSIM(A_m, A_n) = \frac{(2\mu_m\mu_n + C_1)(2\sigma_{mn} + C_2)}{(\mu_m^2 + \mu_n^2 + C_1)(\sigma_m^2 + \sigma_n^2 + C_2)} \quad (11)$$

Due to the inclusion of a consistency measurement in terms of luminance, contrast and structural similarity, SSIM is found to produce more consistent similarity measurements than distance and mutual information based approaches [9]. For the 92AV3C dataset as described in detail in the next section, SSIM-based band similarity maps are shown in Fig. 2. As can be seen, bands are naturally divided into similar groups, with exceptions to bands just over 100 and 150 where adjacent bands show low similarity to each other. These are noisy bands which contain little useful information, hence the low correlation with adjacent bands.



[Fig2] The correlation matrix among the band images of the 92AV3C dataset.

Based on the similarity maps above, band groups and the key bands can be easily determined by thresholding [3] or Jeffries-Matusita interclass distance [7] with the steepest ascent search strategy [2, 7]. For machine learning based approaches, ground-truth information is usually used to determine the discriminatory ability of bands in any classification based tests.

## PERFORMANCE EVALUATION AND ANALYSIS

We compare the performance of SPCA and Seg-PCA with conventional PCA using the publicly available hyperspectral dataset 92AV3C from the original Indian Pines [10]. The 92AV3C dataset was collected by the AVIRIS instrument over an agricultural study site in Northwest Indiana [10] for land cover classification in remote sensing applications. The hyperspectral dataset contains 220 spectral reflectance bands in the wavelength range of 400nm to 2500nm, and the spatial size is 21025 (145×145) pixels. The dataset contains different land cover classes corresponding to agriculture, forest, vegetation, and buildings. One sample band and the ground truth of the 16 land cover classes are illustrated in Fig. 3. Disregarding the white regions which are unclassified background, the other coloured regions in the ground truth refer to alfalfa, corn (3 types), wheat, grass (3 types), oats, woods, soybeans (3 types), hay, and two others.

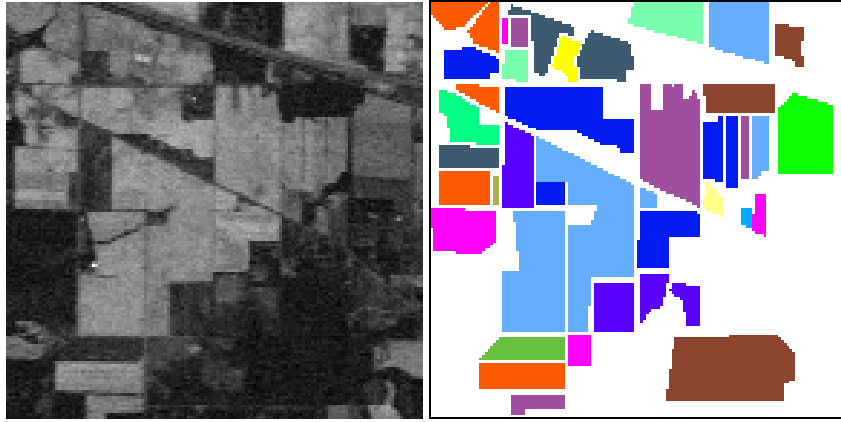
## FEATURE EXTRACTION AND DATA CLASSIFICATION

Support Vector Machine (SVM) has been widely used for hyperspectral image classification [6-7]. The implementation using the bSVM library [11] is employed for performance assessment. As a supervised classifier, SVM relies on training data for model optimization. In the experiments, the principal components extracted from the original spectral data using the conventional PCA, SPCA and Seg-PCA approaches are used as features for the SVM. For 21025 pixels in the dataset, 30% of them are used for training the SVM and the remaining 70% for testing. The kernel function used for

SVM is the Radial Basis Function (RBF), whose parameters, the penalty ( $c$ ) and the gamma ( $\gamma$ ), were optimized in the training process through a grid search of possible combinations of these parameters.

As suggested in the literature [5, 7], the noisy bands covering the region of water absorption (104-108, 150-163, 220) are removed from the spectral vectors before applying PCA for feature extraction. To this end, the effective number of spectral bands for the datasets is reduced from 220 to 200. In total, four groups of features are used for comparison. The first is referred to as the Whole Spectral Band (WSB), in which the total 200 spectral bands are used directly as features. This is taken as a baseline approach for benchmarking. The remaining three groups use principal components as features, extracted from the conventional PCA, the SPCA and Seg-PCA approaches respectively.

For each group of features as samples, 10 experiments were carried out with data samples randomly selected for training and testing. No data sample overlap was allowed in the training set and testing set. The average testing result over these tests and the corresponding standard deviation of classification accuracy were obtained and reported below for evaluation and assessment.



[Fig 3] One sample band image (left) and the ground truth of land cover (right) of the 92AV3C dataset.

## PERFORMANCE ASSESSMENT

We assess performance in terms of reduction of computation cost, memory requirement and improvement of classification accuracy.

The computation costs of the SPCA and Seg-PCA are compared against those of the conventional PCA in Table I, for the three stages involved, where  $S$  and  $K$  refer respectively to the number of samples (pixels) and the number of principal components extracted. Since the computational cost for PCA and SPCA is the same, they are put together to be compared with Seg-PCA. The saving factor of computational cost for Seg-PCA is  $H$  or  $H^3$  in comparison to conventional PCA and SPCA. With  $H = 10$  and  $K = 30$ , the number of Multiply-ACcumulates (MACs) required in PCA/SPCA is  $9.75e8$ , where in Seg-PCA it is reduced to  $9.67e7$ . In effect, the computational cost has reduced to 9.92%. In other words, Seg-PCA has achieved a saving factor of 10, i.e. an order of magnitude for the hyperspectral remote sensing dataset used.

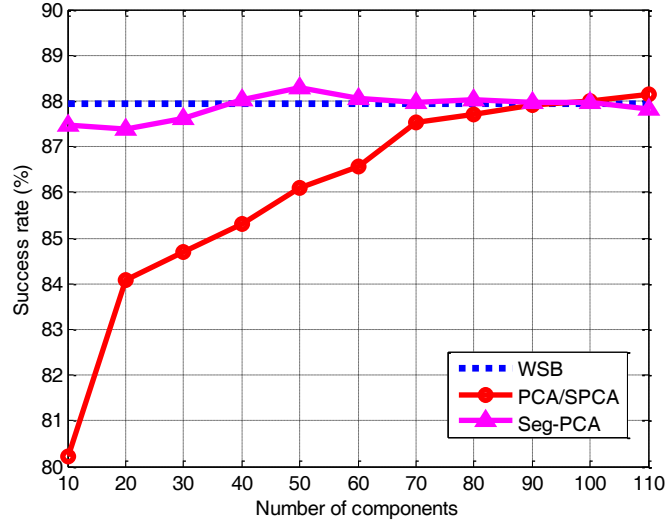
TABLE I. COMPARISONS OF COMPUTATIONAL COST

Stages \ Approaches	Covariance matrix	Eigen problem	Data projection
PCA/SPCA	$O(S H^2 W^2)$	$O(H^3 W^3)$	$O(S HWK)$
Seg-PCA	$O(S H W^2)$	$O(HW^3)$	$O(S WK)$
<i>Saving factor</i>	$H$	$H^2$	$H$

In Table II, memory requirements for PCA, SPCA and Seg-PCA over different stages are compared. SPCA only reduces memory requirement in the first stage when the covariance matrix is obtained. The other two stages require the same amount of memory as the conventional PCA. Seg-PCA has significantly lower memory requirements, and the minimum saving factor achieved is  $H^2$ . When  $H = 10$ , Seg-PCA only requires 1% of the memory compared to the conventional PCA.

TABLE II. COMPARISON OF MEMORY REQUIREMENTS IN PCA, SPCA AND SEG-PCA AT DIFFERENT STAGES

Stages \ Approaches	Covariance computation	Covariance matrix size	Projection matrix size	Projection multiplication
PCA	$S \times HW$	$HW \times HW$	$HW \times K$	$SHW \times HWK$
SPCA	$HW \times 1$	$HW \times HW$	$HW \times K$	$SHW \times HW$
<i>Saving factor</i>	$S$	$1$	$1$	$1$
Seg-PCA	$S \times W$	$W \times W$	$W \times K/H$	$HW \times WK/H$
<i>Saving factor</i>	$H$	$H^2$	$H^2$	$SH^2$



[Fig 4] Classification results for the 92AV3C dataset from PCA/SPCA and Seg-PCA in benchmarking with WSB approach under various numbers of principal components used, where the training ratio is 30%.

With local structures extracted over the spectral domain, Seg-PCA has great potential to improve the efficacy in feature extraction resulting in higher discrimination power and better classification. With  $H = 10$  ( $W = 20$ ), the classification results using various numbers of principal components are plotted in Fig. 4, where  $K$  covers a large range from 10 to 110. First, Seg-PCA is found to consistently outperform conventional PCA (and also SPCA). As can be seen, Seg-PCA significantly improves on the conventional PCA when the number of components is less than 110, i.e. 50% of the dimensionality of original features. When  $K$  is 10 and over, Seg-PCA achieves comparable or even slightly better results in comparison to WSB for classification, a significant advantage for feature extraction as the dimensionality of features used in WSB is 200.

### Non-PCA based approaches

The mutual information based band selection approach [5], SVM based approach [6] and the steepest ascent search based approach [7] are compared in this group of experiments, along with the PCA



based methods. Again SVM is used as the preferred classifier yet under 60% training ratio. The results are summarised in Table III for comparison. As one can see, PCA based approaches have generated reasonably good results, where Seg-PCA is found to be the best among the five approaches evaluated. SVM based approach outperforms conventional PCA and the steepest ascent search based approach, and mutual information based approach is found to be the poorest performing in this group of tests.

TABLE III. COMPARISON OF THE OVERALL ACCURACY OF DIFFERENT APPROACHES UNDER VARIOUS NUMBER OF FEATURES USED FOR CLASSIFICATION

Methods	PCA	Seg-PCA [4]	Mutual information [5]	SVM based band selection [6]	Steepest ascent search method [7]
10	86.76	91.40	73.06	87.55	85.79
20	89.56	92.49	73.95	90.68	87.92
30	89.17	92.61	76.58	91.03	89.45

### Effect of Noise

Due to atmospheric water absorption and other effects, the hyperspectral images obtained may contain severe noise, where the corresponding band image is effectively useless as it has no correlation to any adjacent bands (see in Fig. 2). As a result, noise removal becomes a very important issue. Without noisy band removal, the classification accuracy achieved for the 92AV3C dataset by PCA with 10 components is only around 70%, in comparison to nearly 87% obtained in Table III. Furthermore, some researchers also apply noise removal in the spatial domain, using the known wavelet shrinkage approach [12]. After removal of the noisy bands, it is found that this pre-processing can further improve the overall classification accuracy by 2-3% [12].

### CONCLUSION

Although PCA has been widely used for feature extraction and data reduction, it suffers from three main drawbacks: high computational cost, large memory requirement and low efficacy in processing large datasets such as HSI. This column analysed two variations of PCA, namely SPCA and Seg-PCA. Seg-PCA can further improve classification accuracy whilst significantly reducing the computational cost and memory requirement, without requiring prior knowledge. There is potential to apply similar feature extraction and data reduction techniques in application areas beyond HSI when analysis of large dimensional datasets is required such as magnetic resonance imaging (MRI) and digital video processing.

### RESOURCES

For hyperspectral remote sensing, a series of sensors have been applied in the world, which include AVIRIS [10], ROSIS [13] and HYDICE [10] as well as many others such as TRWIS, CASI, OKSI AVS, MERIS, Hyperio, HICO, CHRIS, NEON and TERN. Detailed specifications in terms of the spectral range and spectral/spatial resolution etc for these sensors are summarised in [14, 15].

For the AVIRIS sensor, the data were acquired since the early 1990s, and they can be obtained from NASA by accessing the link ([http://aviris.jpl.nasa.gov/data/get\\_aviris\\_data.html](http://aviris.jpl.nasa.gov/data/get_aviris_data.html)). Alternatively, some associated datasets such as 92AV3C, Salinas, and Cuprite can be downloaded online from [10] and/or [13]. For the ROSIS sensor, the Pavia Center and Pavia University datasets can be obtained from [13]. In addition, the DC Mall dataset from the HYDICE sensor is also available online [10]. Note that for 92AV3C, Salinas, the Pavia Center and the Pavia University datasets, there are available ground truth to facilitate objective assessment in data classification based applications.

## AUTHORS

*Jinchang Ren* (Jinchang.Ren@strath.ac.uk) is the Deputy Director of Hyperspectral Imaging Centre in the University of Strathclyde, Glasgow, U.K.

*Jaime Zabalza* (J.Zabalza@strath.ac.uk) is currently working towards his PhD. at the Department of Electronic and Electrical Engineering, University of Strathclyde.

*Stephen Marshall* (Stephen.Marshall@strath.ac.uk) is a Professor with the Department of Electronic and Electrical Engineering in Strathclyde, and a Fellow of the IET.

*Jiangbin Zheng* (zhengjb@nwpu.edu.cn) is currently a Professor in School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China.

## REFERENCES

- [1] R. Dianat and S. Kasaei, "Dimension reduction of optical remote sensing images via minimum change rate deviation method," *IEEE Trans. Geosci. Remote Sens.*, 48(1): 198-206, Jan. 2010.
- [2] X. Jia, B.-C. Kuo, M. Crawford, "Feature mining for hyperspectral image classification," *Proc. of IEEE*, vol. 101, no. 3, pp. 676-697, March 2013.
- [3] J. Yang, D. Zhang, A. F. Frangi, and J-Y Yang, "Two-Dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. PAMI*, vol. 26, no.1, pp. 131-137, January 2004
- [4] X. Jia, and J.A. Richards, "Segmented principal components transformation for efficient hyperspectral remote sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, pp. 538-542, 1999.
- [5] B. Guo, S. Gunn, R. Damper, and J. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geoscience and Remote Sensing Letters*, Vol. 3, no. 4, pp. 522-526, 2006
- [6] M. Pal, and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 48, no. 5, pp. 2297-2307, 2010
- [7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with SVMs," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, August 2004.
- [8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Somoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, no. 4, pp. 600-612, 2004
- [9] D. P. Widemann, *Dimensionality Reduction for Hyperspectral Data*, PhD thesis, University of Maryland, 2008.
- [10] Pursue's univeristy multispec site: June 12, 1992 aviris image north-south flightline, indiana [Online], available : <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>
- [11] bSVM multiclass classifier 2.08, released on June 18, 2012 [Online], available: <http://www.csie.ntu.edu.tw/~cjlin/bsvm/>
- [12] B. Demir, S. Erturk, and M. K. Gullu, "Hyperspectral image classification using denoising of intrinsic mode functions," *IEEE Geosc. & Remote Sens. Letters*, vol. 8, no. 2, pp. 220-224, 2011
- [13] Online datasets: [http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)
- [14] [http://www.gisresources.com/fundamemntals-of-hyperspectral-remote-sensing\\_2/](http://www.gisresources.com/fundamemntals-of-hyperspectral-remote-sensing_2/)
- [15] <http://hyspeedblog.wordpress.com/2013/04/23/remote-sensing-data-access-a-review-of-online-resources-for-hyperspectral-imagery/>